

PATENT
Docket No.: 16869N-093500US
Client Ref. No.: NT1252US

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re: application of:

Katsuya Tanaka

Application No.: 10/656,492

Filed: September 5, 2003

For: Disk Array System and Failure
Recovering Control Method

Customer No.: 20350

Confirmation No. 8350

Examiner: Unassigned

Technology Center/Art Unit: 2655

PETITION TO MAKE SPECIAL FOR
NEW APPLICATION PURSUANT TO
37 C.F.R. § 1.102(d) &
M.P.E.P. § 708.02, Item VIII,
ACCELERATED EXAMINATION

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Sir:

This is a petition to make special the above-identified application in accordance with MPEP § 708.02, Item VIII, accelerated examination. The application has not received any examination by the Examiner.

(A) The Commissioner is authorized to charge the petition fee of \$130 under 37 C.F.R. § 1.17(h), and any additional fees that may be associated with this petition may be charged to Deposit Account No. 20-1430.

(B) All the claims are believed to be directed to a single invention. If the examiner determines that all the claims presented are not obviously directed to a single invention, then Applicant will make an election without traverse as a prerequisite to the grant of special status where the specific grouping of claims will be determined by the examiner.

09/22/2005 MAHME1 00000001 201430 10656492

01 FC:1464 130.00 DA

(C) A pre-examination search was performed by an independent patent search firm. The pre-examination search includes a classification search, a computer database search, and a keyword search. The classification search covered the following classes and sub-classes:

711/ 114, 148, 154, 170

714/ 5, 6, 7

Additionally, a computer database search was conducted on the USPTO systems EAST and WEST. The following references were identified in the search report:

- (1) U.S. Patent No.:
5,077,736 Dunphy, Jr. et al.
5,737,510 Kakuta et al.
6,058,489 Schultz et al.
- (2) U.S. Patent Application Publication Nos.:
2002/0174296 Ulrich et al.
2005/0108475 Yamamoto
- (3) Japanese Application Publication No.:
JP-09016343A Kyoichi et al.

(D) The above references are enclosed herewith, collectively as Exhibit A.

(E) Set forth below is a detailed discussion of the references, pointing out with particularity how the claimed subject matter recited in the claims, amended according to the preliminary amendment filed herewith, is distinguishable over the references.

Claimed Subject Matter of the Present Invention

There are two independent claims among the nine pending claims.

Independent claim 1 recites a disk array system having a maintenance terminal connected its disk controller. A disk array is connected to the disk controller via a disk channel. The disk array comprises a plurality of drive boards each mounting thereon a plurality of disk drives connected to wiring on the board. The disk array further comprises a common board provided with a plurality of connectors for connecting the wiring on each of the drive boards to wiring for the disk channel, the a plurality of drive boards being detachably mounted on the

common board via the plurality of connectors. The disk controller, among other things, defines a part of the drive boards in the disk array as spare boards and the rest as active boards, manages $(N + 1)$ pieces of the disk drives ($N \geq 2$) that are mounted on different drive boards in a group of active drive boards as a logical group, and either dispersively allocates a memory area for storing error correction information generated in each logical group to the $(N + 1)$ pieces of disk drivers or fixedly allocates the memory area to a specific disk drive. The disk controller comprises means for reorganizing logical groups when failure occurs in any active disk drive in the disk array, after storing the data that is stored on the disk drives on the faulty board into corresponding disk drives on the substitution board selected from among a group of spare drive boards, so that each logical group to which one of disk drives on the faulty board belongs includes a new disk drive on the substitution board in place of the disk drive on the faulty board. The disk drive further comprises means for informing the maintenance terminal that the faulty board is replaceable after the reorganization of the logical groups is completed.

Independent claim 7 recites a failure recovering control method executed by a disk controller in a disk array system, where a disk array that is connected to the disk controller includes a plurality of drive boards each mounting thereon a plurality of disk drives connected to wiring on the board, and a common board provided with a plurality of connectors for connecting the wiring on each of the drive boards to wiring for the disk channel, the plurality of drive boards being detachably mounted on the common board via the plurality of connectors. The disk controller, among other things, defines a part of the drive boards in the disk array as spare boards and the rest as active boards, manages $(N + 1)$ pieces of disk drives ($N \geq 2$) mounted on different drive boards in a group of active drive boards as a logical group, dispersively allocates a memory area for storing error correction information generated in each logical group to the plurality of disk drives or fixedly allocates the memory area to a specific disk drive. The recited method includes selecting a substitution board to be used in place of a faulty board on which a faulty drive is mounted from among a group of the spare drive boards. The recited method further includes storing the data that is stored in each disk drive on the faulty board into disk drives on the substitution board. The recited method further includes reorganizing logical groups to each of which a disk drive on the faulty board belongs into new configuration including a new

disk drives on the substitution board in place of the disk drive on the faulty board. The recited method further includes informing the maintenance terminal that the faulty board can be replaced.

U.S. Patent No. 5,077,736 Dunphy, Jr. et al.

The patent to Dunphy, Jr. et al. (5,077,736), assigned to Storage Technology Corporation, provides for a Disk Drive Memory. Disclosed is disk drive memory 100 including disc drive manager 140 and having a large plurality of small form factor disk drives 130-0, 130-1 ... 130-M including a pool of backup replacement disk drives, where the drives are organized in redundancy groups and the backup disk drives automatically substitute a failed drive in a redundancy group (see figure 1; and column 9, lines 19-58).

As to **claim 1**, the reference does not show or suggest a disk array comprising a plurality of drive boards each mounting thereon a plurality of disk drives and disk controller that defines a part of the drive boards in the disk array as spare boards and the rest as active boards. The reference does not show or suggest a disk controller storing the data that is stored on the disk drives on a faulty board into corresponding disk drives on a substitution board selected from among the spare drive boards. The reference does not show or suggest means for reorganizing logical groups after the data is stored on drives of the substitution board. The reference does not show or suggest the disk controller informing a maintenance terminal that the faulty board is replaceable after the reorganization of the logical groups is completed.

As to **claim 7**, the reference does not show or suggest failure recovering control method in a disk array system comprising a disk array is connected to a disk controller and includes a plurality of drive boards each mounting thereon a plurality of disk drives where the disk controller defines a part of the drive boards in the disk array as spare boards and the rest as active boards. The reference does not show or suggest a step of selecting a substitution board to be used in place of a faulty board on which a faulty drive is mounted from among a group of the spare drive boards. The reference does not show or suggest a step of storing the data that is stored in each disk drive on the faulty board into disk drives on the substitution board, and then reorganizing logical groups to each of which a disk drive on the faulty board belongs into new

configuration including the new disk drives on the substitution board in place of the disk drive on the faulty board. The reference does not show or suggest a step of informing the maintenance terminal that the faulty board can be replaced.

U.S. Patent No. 5,737,510 Kakuta et al.

The patent to Kakuta et al. (5,737,510), assigned to Hitachi, Ltd., provides for a Disc Array System Having Disc Storage Devices Dispersed on Plural Boards and Accessible at Withdrawal of Part of the Boards. Disclosed is a disc array system including array controller 2 connected through channel path switch 11 to CPU 1, and having drives S mounted in a distributed manner on boards SA, where drives in a parity group are on different boards, and where a board carrying a faulty drive is withdrawn from mother board 37 when a fault occurs on a drive on that board (see figure I; and column 5, lines 9-28).

As to **claim 1**, the reference does not show or suggest a disk array comprising a plurality of drive boards each mounting thereon a plurality of disk drives and disk controller that defines a part of the drive boards in the disk array as spare boards and the rest as active boards. The reference does not show or suggest a disk controller storing the data that is stored on the disk drives on a faulty board into corresponding disk drives on a substitution board selected from among the spare drive boards. The reference does not show or suggest means for reorganizing logical groups after the data is stored on drives of the substitution board. The reference does not show or suggest the disk controller informing a maintenance terminal that the faulty board is replaceable after the reorganization of the logical groups is completed.

As to **claim 7**, the reference does not show or suggest failure recovering control method in a disk array system comprising a disk array is connected to a disk controller and includes a plurality of drive boards each mounting thereon a plurality of disk drives where the disk controller defines a part of the drive boards in the disk array as spare boards and the rest as active boards. The reference does not show or suggest a step of selecting a substitution board to be used in place of a faulty board on which a faulty drive is mounted from among a group of the spare drive boards. The reference does not show or suggest a step of storing the data that is stored in each disk drive on the faulty board into disk drives on the substitution board, and then

reorganizing logical groups to each of which a disk drive on the faulty board belongs into new configuration including the new disk drives on the substitution board in place of the disk drive on the faulty board. The reference does not show or suggest a step of informing the maintenance terminal that the faulty board can be replaced.

U.S. Patent No. 6,058,489 Schultz et al.

The patent to Schultz et al. (6,058,489), assigned to Compaq Computer Corporation, provides for an On-Line Disk Array Reconfiguration. Disclosed is a system including disk controller 10 connected to host computer 101 and having hard disk subsystems 142 including disk drives 114A-E forming a RAID array, where disk array reconfiguration takes place, for example, when a drive is to be removed (see figure 1; and column 2, line 65-column 3, line 3).

As to **claim 1**, the reference does not show or suggest a disk array comprising a plurality of drive boards each mounting thereon a plurality of disk drives and disk controller that defines a part of the drive boards in the disk array as spare boards and the rest as active boards. The reference does not show or suggest a disk controller storing the data that is stored on the disk drives on a faulty board into corresponding disk drives on a substitution board selected from among the spare drive boards. The reference does not show or suggest means for reorganizing logical groups after the data is stored on drives of the substitution board. The reference does not show or suggest the disk controller informing a maintenance terminal that the faulty board is replaceable after the reorganization of the logical groups is completed.

As to **claim 7**, the reference does not show or suggest failure recovering control method in a disk array system comprising a disk array is connected to a disk controller and includes a plurality of drive boards each mounting thereon a plurality of disk drives where the disk controller defines a part of the drive boards in the disk array as spare boards and the rest as active boards. The reference does not show or suggest a step of selecting a substitution board to be used in place of a faulty board on which a faulty drive is mounted from among a group of the spare drive boards. The reference does not show or suggest a step of storing the data that is stored in each disk drive on the faulty board into disk drives on the substitution board, and then

reorganizing logical groups to each of which a disk drive on the faulty board belongs into new configuration including the new disk drives on the substitution board in place of the disk drive on the faulty board. The reference does not show or suggest a step of informing the maintenance terminal that the faulty board can be replaced.

U.S. Patent Application Publication No. 2002/0174296 Ulrich et al.

The patent application publication to Ulrich et al. (2002/0174296 AI) provides for a Disk Replacement via Hot Swapping with Variable Parity. Disclosed is a system including server 130 and disk array 140, where disk array reconfiguration takes place when a drive fails, such as in one embodiment by reconstructing the parity group and storing it in a space newly allocated for a new parity group (see figure 30; and paragraphs 28, 126, 168 and 455).

As to **claim 1**, the reference does not show or suggest a disk array comprising a plurality of drive boards each mounting thereon a plurality of disk drives and disk controller that defines a part of the drive boards in the disk array as spare boards and the rest as active boards. The reference does not show or suggest a disk controller storing the data that is stored on the disk drives on a faulty board into corresponding disk drives on a substitution board selected from among the spare drive boards. The reference does not show or suggest means for reorganizing logical groups after the data is stored on drives of the substitution board. The reference does not show or suggest the disk controller informing a maintenance terminal that the faulty board is replaceable after the reorganization of the logical groups is completed.

As to **claim 7**, the reference does not show or suggest failure recovering control method in a disk array system comprising a disk array is connected to a disk controller and includes a plurality of drive boards each mounting thereon a plurality of disk drives where the disk controller defines a part of the drive boards in the disk array as spare boards and the rest as active boards. The reference does not show or suggest a step of selecting a substitution board to be used in place of a faulty board on which a faulty drive is mounted from among a group of the spare drive boards. The reference does not show or suggest a step of storing the data that is stored in each disk drive on the faulty board into disk drives on the substitution board, and then reorganizing logical groups to each of which a disk drive on the faulty board belongs into new

configuration including the new disk drives on the substitution board in place of the disk drive on the faulty board. The reference does not show or suggest a step of informing the maintenance terminal that the faulty board can be replaced.

U.S. Patent Application Publication No. 2005/0108475 Yamamoto

The patent application publication to Yamamoto (2005/0108475 AI) provides for a Data Save in Disk Array System. Disclosed is disk array system 100 including disk controller 105 connected to host computer 109 and having parity groups 102 including disk drives 101, where, when a physical drive fails, data of the physical drive is saved in a spare logical drive (see figure 3; and paragraph 45).

As to **claim 1**, the reference does not show or suggest a disk array comprising a plurality of drive boards each mounting thereon a plurality of disk drives and disk controller that defines a part of the drive boards in the disk array as spare boards and the rest as active boards. The reference does not show or suggest a disk controller storing the data that is stored on the disk drives on a faulty board into corresponding disk drives on a substitution board selected from among the spare drive boards. The reference does not show or suggest means for reorganizing logical groups after the data is stored on drives of the substitution board. The reference does not show or suggest the disk controller informing a maintenance terminal that the faulty board is replaceable after the reorganization of the logical groups is completed.

As to **claim 7**, the reference does not show or suggest failure recovering control method in a disk array system comprising a disk array is connected to a disk controller and includes a plurality of drive boards each mounting thereon a plurality of disk drives where the disk controller defines a part of the drive boards in the disk array as spare boards and the rest as active boards. The reference does not show or suggest a step of selecting a substitution board to be used in place of a faulty board on which a faulty drive is mounted from among a group of the spare drive boards. The reference does not show or suggest a step of storing the data that is stored in each disk drive on the faulty board into disk drives on the substitution board, and then reorganizing logical groups to each of which a disk drive on the faulty board belongs into new configuration including the new disk drives on the substitution board in place of the disk drive on

the faulty board. The reference does not show or suggest a step of informing the maintenance terminal that the faulty board can be replaced.

Japanese Application No. JP-09016343A Kyoichi et al.

The Official Gazette of JP 09016343 relates to a RAID type disk drive system. A translation of relevant portions (identified by paragraph numbers) of this reference made by the Applicant follows:

(0027) FIG.1 is a diagram showing the configuration of a disk array subsystem of an embodiment of the invention.

(0028) In the Figure, numeral 1 denotes a host apparatus, for example, a computer. Numeral 2 denotes a disk controller, which is a controller of the disk array provided with a RAID function. Numeral 3 denotes a processor for controlling the function of the disc controller 2. Numeral 4 denotes a memory to be used by the processor 3. Numeral 5 denotes a parity control circuit for performing control to generate parity data according to the RAID function, write the parity data into a disk drive (hereafter, referred as "parity disk drives") such as 10d1-10d3 within an array group, read out parity data from a parity disk drive such as 10d1-10d3 when a failure occurs in a disk drive (faulty disk drive) in each array group, recover data in the faulty disc drive based on the data and write the data into a spare disk such as 10e1-10e3. Numeral 6 denotes a buffer control circuit for controlling a buffer memory. Numeral 7 denotes a data buffer memory, which is a buffer memory to store data read out from the disk drives temporally. Numeral 8 denotes a disk interface (hereafter, an interface is referred as I/F) to control an I/F at the disk drive side. Numeral 9 denotes a host I/F to control an I/F at the host apparatus side. Numerals 10a1 to 10e3 denote disk drives, from/to which bit data are read/write through SCSI I/Fs 12 by the disk controller 2. Into disk drives 10d1-10d3, parity data of respective group are recorded. Numeral 11 denotes a SCSI I/F at the host apparatus side as an I/F of the host apparatus 1. Numeral 12 denotes SCSI I/Fs at the disk drive side. The disk controller 2 is provided with a plurality of SCSI I/Fs 12, the number of which corresponds to the number of disk drives within one array group. Numerals 13a to 13e denote a plurality of units each to be a unit of replacement, each of which is constructed to accommodate a predetermined number of disk drives, for example, three disk drives 10#1 to 10#3. A block of replacement unit is each vertical array (one unit). Numerals 14a to 14c denote array groups, each of which is defined by an array, a horizontal array, different from that for the replacement unit. Each array group is a logical group for generating parities according to the RAID function. Numeral 15 denotes an internal system bus within the disk controller 2.

(0037) As shown in FIG. 2, in this disk array subsystem, when a failure occurs in a disk drive, for example, in a disk drive 10c3, the disk controller 2 reads out the contents (bit data such as "1" and "0") from disk drives 10a3, 10b3 and 10d3, other

than the faulty disk drive 10c3, that belong to the same array group 14c (a group of disk drives correlated for generating parity ties) with the faulty disk drive 10c3. The parity control circuit 5 recovers the contents (data) of the faulty drive by calculating Exclusive OR between bit data of a disk drives 10a3 and 10b3 and comparing the result of XOR with parity data of disk drive 10b3, and the buffer control circuit 6 writes the data into a data buffer memory 7.

(0038) Consecutively, the disk controller 2 writes the data written into the data buffer memory 7 into a disk drive (spare disk drive) 10e3, which is on the spare unit 13e and corresponding to the faulty disk drive 10c3.

(0039) Next, the disk controller 2 copies the contents of the other disk drives 10c1 and 10c2 on the same unit 13c with the faulty disk drive 10c3 into corresponding disk drives 10e1 and 10e2 on the spare unit 13e, respectively. The disk controller 2 carries out above processing for data recovering and data copy during when access requests from the host apparatus 1 are absent.

(0040) Upon completing these recovering and copying, the function of the unit 13c including the faulty disk drive 10c3 can be replaced by the spare unit 13e entirely so that the unit 13 of the faulty disk becomes out of use.

(0041) Accordingly, if the unit 13c accommodating the faulty disk drive 10c3 is detached from the SCSI I/F 12 in order to replace the faulty disk drive 10c3 with a new one, since the spare unit 13e stands proxy for the function of the unit 13c, the entire system can operate in the same way as the disk drive 10c3 was in a state before the failure, whereby the system can be operated without exerting an influence on its performance.

(0045) As this manner, according to the embodiment of the disk array subsystem, more than two disk drives 10a1-10a3 are unitized to connect to a SCSI I/F 12, a plurality of units are disposed in a array, and one of the plurality of units 13a-13e is designated as a spare unit 13e. When a certain one of the plurality of disk drives, for example, a disk drive 13c3 in a unit 13c becomes out of order, after the data of faulty disk drive 13c3 are recovered to write into a corresponding spare disk drive 10e3, the data of the other disk drives 10c1 and 10c2 on the same unit 13c with the faulty disk drive 10c3 are copied into disk drives 10e1 and 10e2 on the spare unit 13e, respectively, and by operating the spare unit 13e for the faulty unit, the faulty disk drive 10c3 can be changeable, together with the unit 13c, without exerting an influence on the performance of the system in operation.

As to **claim 1**, the reference does not show or suggest a disk array comprising a plurality of drive boards each mounting thereon a plurality of disk drives and disk controller that defines a part of the drive boards in the disk array as spare boards and the rest as active boards. The reference does not show or suggest a disk controller storing the data that is stored on the disk drives on a faulty board into corresponding disk drives on a substitution board selected from among the spare drive boards. The reference does not show or suggest means for reorganizing logical groups after the data is stored on drives of the substitution board. The reference does not


show or suggest the disk controller informing a maintenance terminal that the faulty board is replaceable after the reorganization of the logical groups is completed.

As to **claim 7**, the reference does not show or suggest failure recovering control method in a disk array system comprising a disk array is connected to a disk controller and includes a plurality of drive boards each mounting thereon a plurality of disk drives where the disk controller defines a part of the drive boards in the disk array as spare boards and the rest as active boards. The reference does not show or suggest a step of selecting a substitution board to be used in place of a faulty board on which a faulty drive is mounted from among a group of the spare drive boards. The reference does not show or suggest a step of storing the data that is stored in each disk drive on the faulty board into disk drives on the substitution board, and then reorganizing logical groups to each of which a disk drive on the faulty board belongs into new configuration including the new disk drives on the substitution board in place of the disk drive on the faulty board. The reference does not show or suggest a step of informing the maintenance terminal that the faulty board can be replaced.

Conclusion

In view of this comments presented in the instant petition and the claim amendments presented in the accompanying preliminary amendment, the Examiner is respectfully requested to issue a first Office Action at an early date.

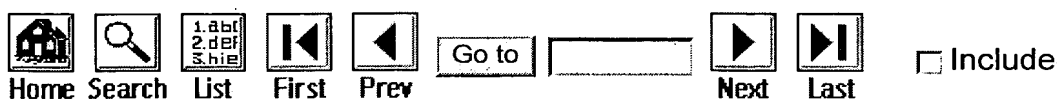
Respectfully submitted,


George B. F. Yee
Reg. No. 37,478

TOWNSEND and TOWNSEND and CREW LLP
Two Embarcadero Center, 8th Floor
San Francisco, California 94111-3834
Tel: 650-326-2400
Fax: 415-576-0300
Attachments
GBFY:cmm
60581818 v1

TABLE OF CONENTS

U.S. Patent No. 5,077,736 to Dunphy, Jr. et al.	1
U.S. Patent No. 5,737,510 to Kakuta et al.	2
U.S. Patent No. 6,058,489 to Schultz et al.....	3
U.S. Patent Application Publication No. 2002/0174296 to Ulrich et al.	4
U.S. Patent Application Publication No. 2005/0108475 to Yamamoto	5
Japanese Application Publication No.: JP-09016343A to Kyoichi et al.	6



MicroPatent® PatSearch Fulltext: Record 1 of 1

Search scope: JP (bibliographic data only)

Years: 1981-2005

Patent/Publication No.: ((JP09016343))

Order This Patent

Family Lookup

Find Similar

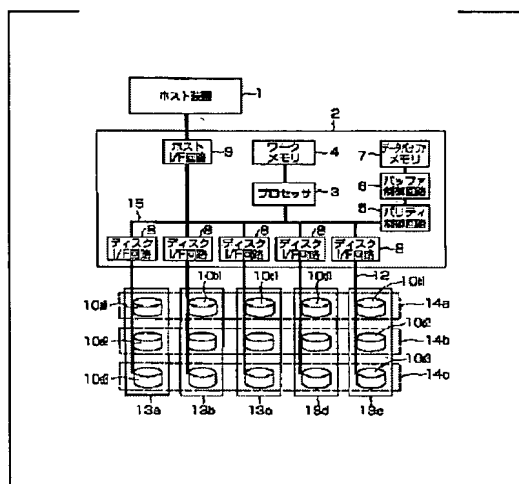
Legal Status

[Go to first matching text](#)

JP09016343 A DISK DRIVE SYSTEM TOSHIBA CORP

Abstract:

PURPOSE: To improve the mounting density of disk drives in the disk array sub-system of a RAID system. **CONSTITUTION:** This disk array sub-system is provided for operating a system by recovering the data of a fault disk drive 10c3 based on the parity data of a parity disk 10d3 when any fault is generated at a certain disk drive 10c3 or the like, disk drives 10a1-10a3 or the like are connected to plural SCSI/F 12 connected with a disk controller 2 three by three and made into a unit and plural exchange units 13a-13e are provided to be exchanged for each unit. On the other hand, when trouble is generated at a certain unit 13c, the disk controller 2 writes the recovered data and the data of the other disk drives 10 in the unit 13c in a spare unit 13e and operates them.



[Click here for larger image.](#)

Inventor(s):

SASAMOTO KYOICHI
KIHARA JUNICHI

Application No. 07166347 JP07166347 JP, **Filed** 19950630, **A1 Published** 19970117

Int'l Class: G06F00306

G06F01216 G11B01902 G11B01904 G11B02010

Patents Citing This One (1):

→ WO9936911 A1 19990722 MATSUSHITA ELECTRIC INDUSTRIAL CO., LTD.
AV DATA INPUT/OUTPUT DEVICE

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平9-16343

(43) 公開日 平成9年(1997)1月17日

(51) Int.Cl. ⁶	識別記号	庁内整理番号	F I	技術表示箇所
G 0 6 F 3/06	5 4 0		G 0 6 F 3/06	5 4 0
	12/16	7623-5B		12/16 3 2 0 L
G 1 1 B 19/02	5 0 1		G 1 1 B 19/02	5 0 1 F
	19/04			19/04 5 0 1 P
	20/10	7736-5D		20/10 D
審査請求 未請求 請求項の数4 O L (全 7 頁)				

(21) 出願番号 特願平7-166347

(22) 出願日 平成7年(1995)6月30日

(71) 出願人 000003078

株式会社東芝

神奈川県川崎市幸区堀川町72番地

(72) 発明者 笹本 享一

東京都府中市東芝町1番地 株式会社東芝

府中工場内

(72) 発明者 木原 淳一

東京都府中市東芝町1番地 株式会社東芝

府中工場内

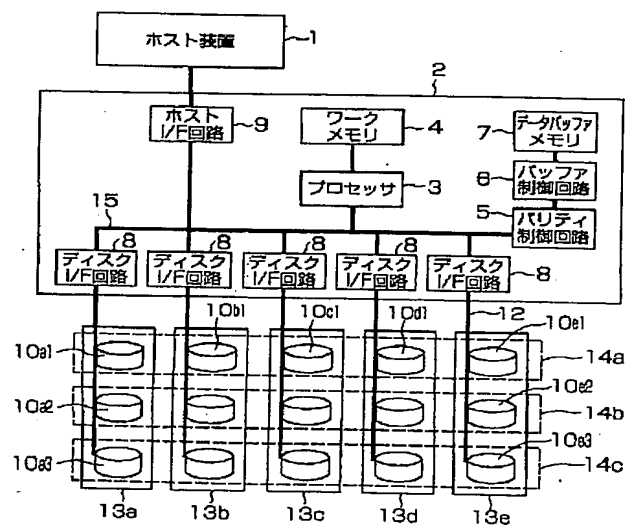
(74) 代理人 弁理士 須山 佐一

(54) 【発明の名称】 ディスクドライブシステム

(57) 【要約】

【目的】 R A I D方式のディスクアレイサブシステムにおいて、ディスクドライブの実装密度を向上する。

【構成】 このディスクアレイサブシステムは、あるディスクドライブ10c3などが故障したときにパリティディスク10d3のパリティデータを基に故障ディスクドライブ10c3のデータを復元してシステムを運用するものであって、ディスクコントローラ2と接続される複数のS C S I I/F 12にそれぞれ3台つづのディスクドライブ10a1~10a3などを接続してユニット化し、そのユニット単位に交換するよう複数の交換ユニット13a~13eを設ける一方、ディスクコントローラ2は、あるユニット13cなどで不具合が発生したとき、復元したデータとユニット13c内の他のディスクドライブ10のデータとをスペアユニット13eに書き込み運用する。



【特許請求の範囲】

【請求項1】 筐体内に複数のディスクドライブと、これらディスクドライブを所定数毎にグループ化し、各グループの中の一つにパリティデータを記録し、故障発生時、前記パリティデータを基に各グループ単位にデータを復元するディスクコントローラとを設けてなるディスクドライブシステムにおいて、

前記各ディスクドライブをグループ間の列毎にまとめて前記ディスクコントローラに対して着脱自在にユニット化してなることを特徴とするディスクドライブシステム。

【請求項2】 請求項1記載のディスクドライブシステムにおいて、

前記ディスクコントローラは、ユニット化した中のあるユニットで不具合が発生したとき、データ復元処理後に前記複数の中の所定ユニットをスペアユニットとして運用することを特徴とするディスクドライブシステム。

【請求項3】 筐体内に複数のディスクドライブと、これらディスクドライブを所定数毎にグループ化し、各グループの中の一つにパリティデータを記録し、故障発生時、前記パリティデータを基に各グループ単位にデータを復元するディスクコントローラとを設けてなるディスクドライブシステムにおいて、

前記所定数毎にグループ化したディスクドライブを列設し、前記ディスクコントローラに対してはグループ間の列毎にまとめて着脱自在に交換ユニットを構成すると共に、

前記ディスクコントローラは、

前記複数の中のある一つのディスクドライブが故障したとき、前記グループ毎に前記パリティディスクのパリティデータを基に他のディスクドライブのデータのパリティチェックを行い、故障ディスクドライブのデータを復元するデータ復元手段と、

前記データ復元手段により復元されたデータと、前記故障ディスクドライブの属するユニット内の他のディスクドライブのデータとを前記複数の中の所定ユニットのそれぞれに対応するディスクドライブに書き込むデータ書込手段と、

前記データ書込手段によりデータの書き込まれた所定ユニットを前記故障ディスクドライブの属するユニットの代替えとしてシステムを運用する制御手段とを具備したことを特徴とするディスクドライブシステム。

【請求項4】 筐体内に複数のディスクドライブと、これらディスクドライブを所定数毎にグループ化し、そのグループ単位にデータ処理を実行するディスクコントローラとを設けてなるディスクドライブシステムにおいて、

前記所定数毎にグループ化したディスクドライブを列設し、前記ディスクコントローラに対してはグループ間の列毎にまとめて着脱自在に交換ユニットを構成すると共

に、

前記ディスクコントローラは、

前記複数のユニットの中の一つをパリティユニットとし、そのパリティユニットに他のユニットのディスクドライブが故障したときのパリティチェック用のパリティデータを記録するパリティデータ記録手段と、

あるユニット内でディスクドライブが故障したときに、前記パリティユニットの対応するディスクドライブから前記パリティデータを読み出し、それを基に故障ディスクドライブのデータを復元するデータ復元手段と、

前記複数のユニットの中の一つをスペアユニットとして、前記データ復元手段により復元された故障ディスクドライブのデータを前記スペアユニットの対応するディスクドライブに書き込む第1の書込手段と、

前記故障ディスクドライブが属するユニット内の他のディスクドライブのデータを前記スペアユニットのそれぞれに対応するディスクドライブに書き込む第2の書込手段と、

前記各データの書き込まれた前記スペアユニットを前記故障ディスクドライブの属するユニットの代替えとしてシステムを運用する制御手段とを具備したことを特徴とするディスクドライブシステム。

【発明の詳細な説明】

【0001】

【産業上の利用分野】本発明は、例えば複数台のディスクドライブを制御するRAID方式などを採用したディスクシステムに関する。

【0002】

【従来の技術】従来から、RAID方式のディスクシステム、例えばディスクアレイサブシステムなどでは、あるディスクドライブが故障したときの交換作業性を考慮して筐体設計をしており、このため筐体形状の制約が多いため、筐体内のディスク実装密度も低く改善が望まれている。

【0003】図4に示すように、従来のディスクアレイサブシステムは、個々のディスクドライブ41a～41eの故障を想定してディスクドライブ1台づつを交換する単位として各ユニット42a～42e内に收容し、それをキャビネット40の表面あるいは裏面に沿って一列（アレイ状）に実装するようにしている。

【0004】この場合、システムが稼働中に、あるユニットのディスクドライブが故障しても、キャビネット40の表面側の扉を開ければ全てのユニット42a～42eが表れ、故障ディスクドライブのユニットを容易に取り外し、新しいものに交換することができる。

【0005】ところで、通常、この種のディスクアレイサブシステムを設計する場合、そのキャビネットの縦、横および奥行きなどの寸法は、安定性などを得る面から予めある程度の寸法に決められている。したがって、上記配置条件でキャビネット40内にディスクドライブの

ユニット42a~42eを配置した後、コントローラや電源装置などのユニット44を配置すると、その内部スペースに空間45ができることになる。

【0006】

【発明が解決しようとする課題】しかしながら、これでは、キャビネット内部の利用効率が悪いばかりか、実装できるディスクドライブ数が非常に少ないため、ディスクドライブの実装効率も悪いという問題がある。

【0007】そこで、ディスクドライブの交換性のある程度無視して、キャビネットの奥行き方向に複数のディスクドライブを配置し、それをコントローラのインタフェースに接続することが考えられるが、この場合、どのディスクドライブが故障してもよいようにキャビネットの周囲を扉だらけにするしかなく、またシステム稼働中（他のディスクドライブが動作しているときなど）にインタフェースから故障ディスクドライブを取り外すと、そのときにノイズが発生し、同一インタフェースに接続されている他のディスクドライブにノイズが入り込み、信号が不安定な状態になるため、通常、適当なノイズ対策（ハード的またはソフト的な対策手段）を付加する必要がある、単にキャビネットの空きスペースにディスクドライブを増設するわけにはいかないという問題があった。

【0008】本発明はこのような課題を解決するためになされたもので、活線挿抜時に他に影響が及ばないようにノイズ対策を追加することなく筐体内のディスク実装効率を向上することのできるディスクシステムを提供することを目的としている。

【0009】

【課題を解決するための手段】上記した目的を達成するために、請求項1記載のディスクドライブシステムは、筐体内に複数のディスクドライブと、これらディスクドライブを所定数毎にグループ化し、各グループの中の一つにパリティデータを記録し、故障発生時、前記パリティデータを基に各グループ単位にデータを復元するディスクコントローラとを設けてなるディスクドライブシステムにおいて、前記各ディスクドライブをグループ間の列毎にまとめて前記ディスクコントローラに対して着脱自在にユニット化してなることを特徴としている。

【0010】また請求項2記載のディスクドライブシステムは、請求項1記載のディスクドライブシステムにおいて、前記ディスクコントローラは、ユニット化した中のあるユニットで不具合が発生したとき、データ復元処理後に前記複数の中の所定ユニットをスペアユニットとして運用することを特徴としている。

【0011】さらに請求項3記載のディスクドライブシステムは、筐体内に複数のディスクドライブと、これらディスクドライブを所定数毎にグループ化し、各グループの中の一つにパリティデータを記録し、故障発生時、前記パリティデータを基に各グループ単位にデータを復

元するディスクコントローラとを設けてなるディスクドライブシステムにおいて、前記所定数毎にグループ化したディスクドライブを列設し、前記ディスクコントローラに対してはグループ間の列毎にまとめて着脱自在に交換ユニットを構成すると共に、前記ディスクコントローラは、前記複数の中のある一つのディスクドライブが故障したとき、前記グループ毎に前記パリティディスクのパリティデータを基に他のディスクドライブのデータの

05 交換ユニットを構成すると共に、前記ディスクコントローラは、前記複数の中のある一つのディスクドライブが故障したとき、前記グループ毎に前記パリティディスクのパリティデータを基に他のディスクドライブのデータの復元するデータ復元手段と、前記データ復元手段により復元されたデータと、前記故障ディスクドライブの属するユニット内の他のディスクドライブのデータとを前記複数の中の所定ユニットのそれぞれ対応するディスクドライブに書き込むデータ書込手段と、前記データ書込手段によりデータの書き込まれた所定ユニットを前記故障ディスクドライブの属するユニットの代替えとしてシステムを運用する制御手段とを具備している。

【0012】また請求項4記載のディスクドライブシステムは、筐体内に複数のディスクドライブと、これらディスクドライブを所定数毎にグループ化し、そのグループ単位にデータ処理を実行するディスクコントローラとを設けてなるディスクドライブシステムにおいて、前記所定数毎にグループ化したディスクドライブを列設し、前記ディスクコントローラに対してはグループ間の列毎にまとめて着脱自在に交換ユニットを構成すると共に、前記ディスクコントローラは、前記複数のユニットの中の一つをパリティユニットとし、そのパリティユニットに他のユニットのディスクドライブが故障したときの

10 パリティチェック用のパリティデータを記録するパリティデータ記録手段と、あるユニット内でディスクドライブが故障したときに、前記パリティユニットの対応するディスクドライブから前記パリティデータを読み出し、それを基に故障ディスクドライブのデータを復元するデータ復元手段と、前記複数のユニットの中の一つをスペアユニットとして、前記データ復元手段により復元された故障ディスクドライブのデータを前記スペアユニットの対応するディスクドライブに書き込む第1の書込手段と、前記故障ディスクドライブが属するユニット内の他のディスクドライブのデータを前記スペアユニットのそれぞれ対応するディスクドライブに書き込む第2の書込手段と、前記各データの書き込まれた前記スペアユニットを前記故障ディスクドライブの属するユニットの代替えとしてシステムを運用する制御手段とを具備している。

45 【0013】

【作用】請求項1記載の発明では、複数のディスクドライブが、所定数毎にグループ化されたものが列設され、各ディスクドライブはディスクコントローラに対してグループ間の列毎にまとめて着脱自在にユニット化されて

50 いる。

【0014】したがって、あるユニット内のディスクドライブが故障した場合に、そのディスクドライブのデータを復元した後、そのディスクドライブが属するユニットの全てのデータを他のユニットへ複写し、故障したディスクドライブの属するユニットをディスクコントローラから取り外せば、ノイズが発生してもユニットごとに取り外すので、同一インターフェイスに接続された他のディスクドライブも取り外されるため、これによる性能の劣化は起こらなくなる。

【0015】また請求項2記載の発明では、ユニット化した中のあるユニットで不具合が発生したとき、データ復元処理後に複数の中の所定ユニットをスペアユニットとして運用するので、不具合が発生したユニットを抜去しても、システム全体の性能が低下することがなくなる。

【0016】さらに請求項3記載の発明では、複数の中のある一つのディスクドライブが故障したとき、ディスクコントローラのデータ復元手段により、グループ毎にパリティディスクのパリティデータを基に他のディスクドライブのデータのバリディチェックが行われて、故障ディスクドライブのデータが復元される。

【0017】そしてこの復元されたデータと、故障ディスクドライブの属するユニット内の他のディスクドライブのデータとがデータ書込手段により複数の中の所定ユニットのそれぞれに対応するディスクドライブに書き込まれる。データが書き込まれると、所定ユニットは故障ディスクドライブの属するユニットの代替えとして制御手段により運用される。

【0018】すなわち、あるユニットで故障が起これば、その内容は全て所定ユニットに再現されて、その機能も全て所定ユニットによって代替される。

【0019】したがって、故障ディスクドライブの属するユニットごとと交換しても、性能劣化がなくなる。

【0020】また請求項4記載の発明では、複数の中のある一つのディスクドライブが故障したとき、ディスクコントローラのデータ復元手段により、グループ毎にパリティディスクのパリティデータを基に他のディスクドライブのデータのバリディチェックが行われて、故障ディスクドライブのデータが復元される。

【0021】すると、まず、復元されたデータが、第1の書込手段によりスペアユニットの対応するディスクドライブに書き込まれる。続いて、第2の書込手段により、故障ディスクドライブが属するユニット内の他のディスクドライブのデータがスペアユニットのそれぞれに対応するディスクドライブに書き込まれる。

【0022】そして各データの書き込まれたスペアユニットは、制御手段により故障ディスクドライブの属するユニットの代替えとしてシステム上で運用される。

【0023】したがって、故障ディスクドライブの属するユニットごとと交換しても、性能劣化が起こらなくな

る。

【0024】上記により、故障ディスクドライブを新しいディスクドライブに交換するためにシステムからそのユニットを取り外してもノイズの影響がなくなり、システム動作中でも他のユニットに影響を与えることなく交換することができる。

【0025】この結果、活線挿抜時に他に影響が及ばないようにノイズ対策を追加することなく筐体内のディスク実装効率を向上することができる。

【0026】

【実施例】以下、本発明の実施例を図面を参照して詳細に説明する。

【0027】図1は本発明に係る一実施例のディスクアレイサブシステムの構成を示す図である。

【0028】同図において、1はホスト装置であり、例えば計算機などである。2はディスクコントローラであり、RAID機能を備えたディスクアレイの制御装置である。3はプロセッサであり、ディスクコントローラ2の機能を制御するものである。4はワークメモリであり、プロセッサ3により利用されるメモリである。5はパリティ制御回路であり、RAID機能に基づきパリティデータを生成すると共に、そのパリティデータをアレイグループ内のあるディスクドライブ（以下パリティディスクドライブと称す）10d1～10d3などに記録し、各アレイグループ内のあるディスクドライブ（故障ディスクドライブ）に故障が発生したとき、パリティディスクドライブ10d1～10d3などからパリティデータを読み出し、そのデータを基に故障ディスクドライブのデータを復元し、スペアディスク10e1～10e3などに書き込むための制御を行う。6はバッファ制御回路であり、バッファメモリを制御するものである。7はデータバッファメモリであり、ディスクドライブから読み出したデータを一時的に格納するバッファメモリである。8はディスクインタフェース回路（以下インタフェースをI/Fと称す）であり、ディスクドライブ側のI/Fを制御するものである。9はホストI/F回路であり、ホスト装置1側のI/Fを制御するものである。10a1～10e3はディスクドライブであり、ディスクコントローラ2によりSCSI I/F 12を通じてビットデータがリード/ライトされる。なおディスクドライブ10d1～10d3には、それぞれのグループのパリティデータが記録される。11はホスト装置側のSCSI I/Fであり、ホスト装置1とのI/Fである。12はディスクドライブ側のSCSI I/Fであり、1つのアレイグループ内のディスクドライブ台数分、ディスクコントローラ2に設けられている。13a～13eは交換単位となる複数のユニットであり、それぞれが所定数のディスクドライブ、例えば3台のディスクドライブ10#1～10#3などを収容して構成されている。交換の単位となるのは縦の列単位（1ユニット）である。14a～14cはアレイグループであり、

交換単位とは異なる列、つまり横の列で決められており、RAID機能によりパリティを生成するディスクドライブの論理的なグループである。15は内部システムバスであり、ディスクコントローラ2内のデータバスである。

【0029】ここで、RAID機能について説明する。

【0030】一般に、RAID機能のレベルとしては、レベル1(RAID1)からレベル5(RAID5)などがある。

【0031】RAID1は、従来より広く採用されているミラーリング(2重化ディスク、ミラードディスクなどと称するもの)である。

【0032】RAID2は、メモリにてしばしば使用されるECCの機能をディスクアレイに適用したものである。複数台のデータHDD(ハードディスクドライブ)に対し、ECCチェックデータとなる複数台のチェックHDDを冗長付加し、ECC機能によりデータを復元する。但し、このレベルは、一般的に実際のソリューションとしては考えられておらず、現在、実現製品はほとんど存在していない。

【0033】RAID3は、ユーザデータをストラッピングにより複数台のHDDに分散して格納する。またこれらHDDグループに対しパリティデータ専用のHDDを付加し、データを復元する機能である。ストラッピングサイズは、1バイトもしくはブロックであり、ユーザデータが複数台のHDDに対してパラレルに格納されると同時にパリティを生成し、これをパリティHDDに格納する。この機能のメリットは、HDDに対してパラレルにアクセスするので高速なデータ転送レートが実現できる。

【0034】RAID4は、RAID3と同様にストラッピング+パリティ専用のHDDを有するが、そのストラッピングサイズが比較的大きく設定された機能である。1回のユーザアクセスは、RAID3がアレイ内のHDDへパラレルにアクセスされるのに対してRAID4では、ストラッピングサイズが大きいためアレイ内の一部のHDDのみにアクセスされる。この機能のメリットは1回のユーザデータアクセスがアレイ内の一部のHDDに対してのみ実施されるため、残りのHDDへのユーザアクセスが可能となり、複数リクエストの同時実行が可能となる。このレベルは一般的に実際のソリューションとしては考えられておらず、現在、実現製品はほとんど存在していない。

【0035】RAID5は、RAID3と同様のストラッピングサイズによるストラッピングを有するが、パリティデータを格納するHDDが固定されずに、アレイ内の各HDDにパリティデータを分散して格納する機能である。この機能のメリットは、RAID4がパリティHDDが固定されているため、そのパリティHDDへのアクセスが性能的なボトルネックとなるのに対し、RAI

D5では、パリティデータが各HDDへ分散されて格納されるため、特定のHDDがボトルネックとなることはない。

【0036】次に図2を参照してこのディスクアレイサブシステムの動作を説明する。

【0037】図2に示すように、このディスクアレイサブシステムでは、ある1台のディスクドライブ、例えばディスクドライブ10c3などが故障した場合、ディスクコントローラ2は、故障したディスクドライブ10c3と同一のアレイグループ14c(パリティを生成する対象となるディスクドライブのグループ)に属する故障したディスクドライブ10c3以外のディスクドライブ10a3、10b3、10d3の内容(“1”、“0”などのビットデータ)を読み出す。そしてパリティ制御回路5にて、ディスクドライブ10a3、10b3のビットデータの排他的論理和を取り、ディスクドライブ10d3のパリティデータと比較することにより、故障したドライブの内容(データ)を復元し、そのデータをバッファ制御回路6がデータバッファメモリ7に書き込む。

【0038】続いて、ディスクコントローラ2は、このデータバッファメモリ7に書き込まれたデータを、故障ディスクドライブ10c3に対応するスペアユニット13e内のディスクドライブ(スペアディスクドライブ)10e3に書き込む。

【0039】次にディスクコントローラ2は、故障ディスクドライブ10c3と同一のユニット13cの他のディスクドライブ10c1、10c2の内容をそれぞれ対応するスペアユニット13e内のディスクドライブ10e1、10e2へコピー(複写)する。なお上記した復元処理ならびにコピー処理などは、ホスト装置1からのアクセス要求がないときにディスクコントローラ2が実行する。

【0040】これらの復元処理と複写処理が完了すると、故障ディスクドライブ10c3を含むユニット13cの機能は、スペアユニット13eにより完全に代替されるようになり故障ディスクのユニット13cは使用されなくなる。

【0041】したがって、故障ディスクドライブ10c3を新しいものに交換するために、故障ディスクドライブ10c3を収容したユニット13cを、SCSI/F12から抜去した場合でも、スペアユニット13eによりその機能が完全に代替されているので、ディスクドライブ10c3が故障する以前の状態と同様にシステム全体が動作し、性能に影響なくシステムを運用することができる。

【0042】ここで、もし従来の制御形態のまま、このように複数のディスクドライブを収容したユニットで交換を実施すると、故障ディスクドライブのデータはスペアのディスクドライブによって代替されるものの、その他のディスクドライブについてはそのまま使用されているため、ユニット抜去後に、スペアに複写されなかった

ディスクドライブ10c1やディスクドライブ10c2に対するデータの復元処理が実施されることになり、この復元処理のためにディスクシステムそのものの性能低下が発生する。

【0043】一方、本実施例では、同一I/Fに接続された複数のディスクドライブを1ユニット化し、システムが稼働状態でも性能および信頼性に影響を与えることなく、複数のディスクドライブをユニットごとの交換可能にし、これによりキャビネットの奥行き方向の空間がディスクドライブの実装用の空間として利用できるようになり、多くのディスクドライブがキャビネット内に実装できるようになる。

【0044】また故障ユニット13cをSCSI I/F 12から抜去したときにノイズが発生しても、同一I/Fに接続されている全てのディスクドライブ10c1～10c3と一緒に取り外してしまうので、そのときのノイズは他に影響を及ぼさなくなり性能劣化が起こることもない。

【0045】このように本実施例のディスクアレイサブシステムによれば、2台以上のディスクドライブ10a1～10a3をSCSI I/F 12に複数接続して1ユニット化し、そのユニットをアレイ状に複数配置し、複数のユニット13a～13e中にスペアユニット13eを設定し、複数の中のあるディスクドライブ、例えばユニット13cのディスクドライブ10c3などが故障したとき、故障したディスクドライブ10c3のデータを復元して対応するスペア用のディスクドライブ10e3に書き込み後、故障ディスクドライブ10c3を格納したユニット13c内の他のディスクドライブ10c1、10c2のデータをスペアユニット13eのそれぞれのディスクドライブ10e1、10e2にコピーし、スペアユニット13eを代替動作させることにより、システム稼働中でも性能に影響することなく故障ディスクドライブ10c3をユニット13cごと交換することが可能となる。

【0046】したがって、図3に示すように、複数のディスクドライブ10a1～10a3をキャビネット20内の奥行き方向に収容したユニット13aをアレイ状に複数列設できるようになり、図4に示した従来のキャビネット40内で無駄な空間となっていた奥行き方向のスペース45をディスクドライブの実装スペースとして有効に使用することができる。

【0047】これによりディスクドライブシステムの実装密度が向上すると共に、装置形状の制約が軽減されるなどの効果が得られる。

【0048】なお上記実施例では、RAID機能レベル3またはレベル4のパリティ構成について説明したが、本発明がその他のレベルについても上記同様に適用できることは言うまでもない。

【0049】

【発明の効果】以上説明したように請求項1記載の発明によれば、ディスクドライブはディスクコントローラに

対してグループ間の列毎にまとめて着脱自在にユニット化されているので、ディスクコントローラからユニットを取り外したときにノイズが発生しても他のディスクドライブへの影響がなくなり性能劣化は起こらなくなる。

また請求項2記載の発明によれば、ユニット化した中のあるユニットで不具合が発生したとき、データ復元処理後にディスクコントローラが複数の中の所定ユニットをスペアユニットとして運用するので、不具合が発生したユニットを抜去しても、システム全体の性能が低下することがなくなる。

【0050】さらに請求項3記載の発明によれば、復元されたデータと、故障ディスクドライブの属するユニット内の他のディスクドライブのデータとがデータ書込手段により複数の中の所定ユニットのそれぞれに対応するディスクドライブに書き込まれて、所定ユニットが故障ディスクドライブの属するユニットの代替えとして運用されるので、故障ディスクドライブの属するユニットごと交換しても、性能劣化が起こらなくなる。

【0051】また請求項4記載の発明によれば、復元されたデータは、まず、第1の書込手段によりスペアユニットの対応するディスクドライブに書き込まれ、続いて第2の書込手段により、故障ディスクドライブが属するユニット内の他のディスクドライブのデータがスペアユニットのそれぞれに対応するディスクドライブに書き込まれて、そのスペアユニットが故障ディスクドライブの属するユニットの代替えとしてシステム上で運用されるので、故障ディスクドライブの属するユニットごと交換しても、性能劣化が起こらなくなる。

【0052】上記により、故障ディスクドライブを新しいものに交換するためにユニット単位で取り外してもノイズの影響がなくなり、システム動作中でも他のユニットに影響を与えることなく交換することができる。

【0053】この結果、活線挿抜時に他に影響が及ばないようにノイズ対策を追加することなく筐体内のディスク実装効率を向上することができる。

【図面の簡単な説明】

【図1】本発明に係る一実施例のディスクアレイサブシステムの構成を示す図である。

【図2】このディスクアレイサブシステムの動作を示す図である。

【図3】このディスクアレイサブシステムにおいて、筐体内に複数のディスクドライブを実装した例を示す図である。

【図4】従来のディスクアレイサブシステムの実装例を示す図である。

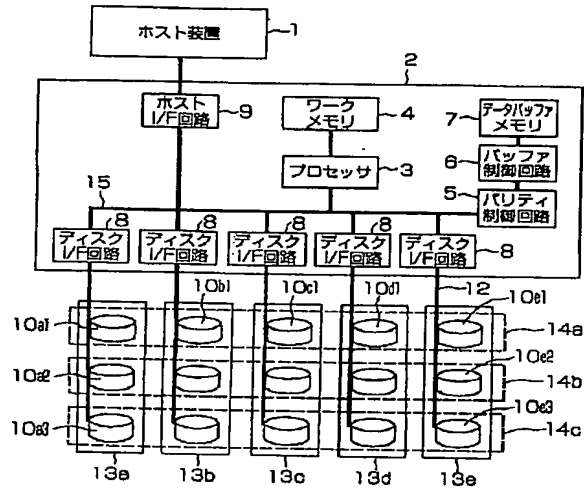
【符号の説明】

1…ホスト装置、2…ディスクコントローラ、3…プロセッサ、4…ワークメモリ、5…パリティ制御回路、6…バッファ制御回路、7…データバッファメモリ、8…ディスクI/F回路、9…ホストI/F回路、10…ディス

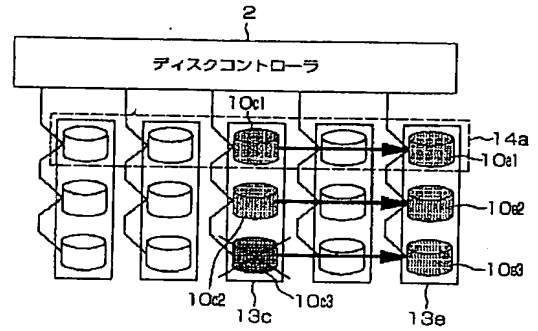
クドライブ、11、12…SCSI I/F、13a~13e…交換単位ユニット、14a~14c…アレイグレー

プ。

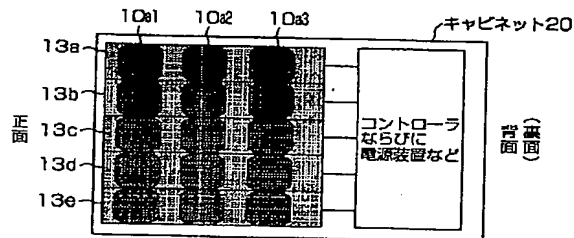
【図1】



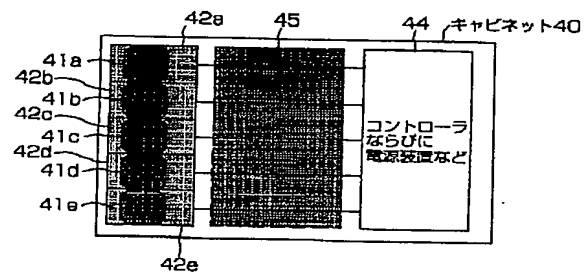
【図2】



【図3】



【図4】



BEST AVAILABLE COPY